



Open Access Repository

www.ssoar.info

Nonresponse Bias (Version 1.1)

Koch, Achim; Blohm, Michael

Erstveröffentlichung / Primary Publication

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Koch, A., & Blohm, M. (2015). *Nonresponse Bias (Version 1.1)*. (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/gesis-sg_004

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

GESIS Survey Guidelines

Nonresponse Bias

Achim Koch & Michael Blohm

Zusammenfassung

Nonresponse ist eine Fehlerquelle in Umfragen. Im vorliegenden Beitrag wird das Konzept des Nonresponse bias erläutert und der Zusammenhang zwischen Response rate und Nonresponse bias diskutiert. Verschiedene Methoden zur Bestimmung von Nonresponse bias werden vorgestellt und es wird erörtert, wie dem Problem des Nonresponse bias bei der Datenerhebung begegnet werden kann.

Zitierung

Koch, Achim und Blohm, Michael (2015). Nonresponse Bias. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_004

*"No issue in survey research is more misunderstood
or controversial than nonresponse."
(Dixon & Tucker, 2010)*

1. Einleitung

Nonresponse (NR) ist eine der möglichen Fehlerquellen in Umfragen. NR tritt auf, wenn es nicht gelingt, für alle Einheiten einer Zufallsstichprobe Daten zu erheben. Traditionell wird danach unterschieden, ob sämtliche Daten für eine Einheit fehlen (Unit nonresponse) oder lediglich ein Teil der Daten nicht gewonnen werden konnte (Item nonresponse). Die vorliegenden Ausführungen gelten dem Problem des Unit nonresponse. Sie beziehen sich auf das Modell einer persönlich-mündlichen Bevölkerungsbefragung, gelten jedoch – ggfls. in etwas modifizierter Weise – auch für andere Befragungsformen.

In der Umfrageforschung ist NR ein allgegenwärtiges Problem, da es praktisch nie gelingt, mit allen ausgewählten Zielpersonen ein Interview durchzuführen. Bei face-to-face Umfragen wird ein Teil der Personen während der Feldzeit nicht erreicht, ein Teil wird zwar kontaktiert, ist aber nicht fähig oder nicht bereit, an der Umfrage teilzunehmen. In vielen westlichen Ländern ist der Anteil dieser Ausfälle in den vergangenen Jahrzehnten gestiegen. Die Ausfälle werfen in mehrerer Hinsicht Probleme auf. Erstens erhöht sich der Stichprobenfehler, wenn der realisierte Stichprobenumfang nicht dem ursprünglich geplanten entspricht. Zweitens sind die Kosten der Umfrage umso höher, je aufwendiger die Prozeduren der Datenerhebung zur Erzielung einer bestimmten Ausschöpfung und Stichprobengröße sind. Drittens besteht aufgrund von Ausfällen die Gefahr einer Verzerrung (Nonresponse bias, NRB), wenn sich die Ausfälle nicht zufällig über die Zielpopulation verteilen.

In der Praxis wird häufig die Response rate (RR), also der Anteil der tatsächlich erzielten Interviews an der ursprünglich ausgewählten Stichprobe, als Proxy für eine Messung des NRB herangezogen. Man nimmt dabei an, dass die Verzerrung umso geringer ist, je höher die Teilnahmequote ausfällt. Dies ist jedoch nur bedingt richtig, wie wir im Folgenden sehen werden. Im nächsten Abschnitt wird der grundlegende Zusammenhang von RR und NRB erläutert. Anschließend werden verschiedene Methoden zur empirischen Bestimmung des NRB vorgestellt. Im vierten Abschnitt wird diskutiert, wie dem Problem des NRB bei der Datenerhebung begegnet werden kann.

2. Response rate und Nonresponse bias

Die RR zählt zu den am häufigsten verwendeten Indikatoren wenn es um die Qualität von Umfragen geht. Die Attraktivität der RR als Qualitätsindikator beruht u.a. auf ihrer (vermeintlich) einfachen Messung und dem Umstand, dass damit ein einzelnes kompaktes Qualitätsmaß für eine Umfrage zur Verfügung steht. Diese Sichtweise lässt allerdings außer Acht, dass Nonresponse nur eine potentielle Fehlerquelle von Umfragen ist (neben Stichprobenfehlern, Messfehlern, etc., vgl. Groves et al., 2009). Außerdem gilt es zu beachten, dass die RR auch für den NRB von Umfragen nur einen unzureichenden Indikator darstellt, da sie lediglich eine Komponente der Stärke des NRB betrifft.

Bei linearen Statistiken wie Mittelwerten oder Anteilswerten ist der NRB eine multiplikative Verknüpfung der Höhe der Nonresponse rate (NRR) einerseits und der Differenz zwischen Teilnehmern und Nichtteilnehmern der Umfrage bei der betrachteten Variable andererseits (vgl. Groves, 2006).

Der NRB eines Mittelwertes \bar{y} z.B. berechnet sich nach

$$NRB(\bar{y}) = NRR * (\bar{y}^T - \bar{y}^{NT})$$

wobei NRR der Nonresponse rate, \bar{y}^T dem Mittelwert der Teilnehmer und \bar{y}^{NT} dem Mittelwert der Nichtteilnehmer bei der betreffenden Surveyvariable entspricht.

Die Formel verdeutlicht zweierlei:

1. Aus der Höhe der NRR kann nur auf das mögliche Potential eines NRB geschlossen werden: je höher die NRR, umso größer kann die Verzerrung ausfallen. Prinzipiell ist es jedoch auch bei einer nur geringen Ausschöpfung möglich, dass der Bias klein ist, nämlich dann, wenn die Ausfälle weitgehend zufällig erfolgen, d.h. wenn die Unterschiede zwischen Teilnehmern und Nichtteilnehmern bei einer Variable gering sind.
2. Der NRB ist eine merkmalspezifische Größe. Eine schlecht ausgeschöpfte Umfrage kann bei einem Merkmal eine deutliche Verzerrung aufweisen (weil sich Teilnehmer und Nichtteilnehmer hinsichtlich der betreffenden Variable unterscheiden), bei einem anderen Merkmal jedoch weitgehend unverzerrt sein (weil Teilnehmer und Nichtteilnehmer bezüglich dieser Variable nicht oder nur wenig differieren). Eine Metaanalyse von NRB-Studien ergab, dass ein großer Teil der Variation von NRB-Schätzern innerhalb von Studien liegt (und nicht zwischen Studien mit unterschiedlicher Ausschöpfung) (Groves, 2006).

Die Aufteilung einer Zufallsstichprobe in die beiden Gruppen der Teilnehmer und Nichtteilnehmer ist ein einfaches deterministisches Modell. In jüngerer Vergangenheit hat ein alternatives stochastisches Modell zunehmend Bedeutung erlangt (Groves, 2006). Es geht davon aus, dass sich Personen nicht einfach entweder als Teilnehmer oder Nichtteilnehmer von Umfragen charakterisieren lassen. Je nachdem, wie eine Umfrage gestaltet wird, resultieren unterschiedliche Aufteilungen derselben Population in Teilnehmer und Nichtteilnehmer. Dies legt nahe, dass jede Einheit einer Stichprobe eine bestimmte, von Null verschiedene Wahrscheinlichkeit hat, an einer Umfrage teilzunehmen (eine sog. 'response propensity') – abhängig von deren Thema, ihrer Länge, der Verwendung von Incentives, dem Verhalten der Interviewer, etc. Diese Wahrscheinlichkeit kann nicht direkt beobachtet, sondern nur geschätzt werden. In diesem Modell lässt sich der NRB des Mittelwertes einer Surveyvariable (y) durch das Verhältnis der Kovarianz des Merkmals mit der response propensity (p) und der mittleren response propensity (welche der RR entspricht) darstellen:

$$NRB(\bar{y}) = \frac{\sigma_{y,p}}{\bar{p}}$$

Nach dieser Formel steigt der Bias mit der Höhe der Kovarianz zwischen der Teilnahmewahrscheinlichkeit und dem betreffenden Merkmal. Sie verdeutlicht, dass ein NRB nur dann auftritt, wenn eine Beziehung zwischen der Teilnahmewahrscheinlichkeit und der interessierenden substantiellen Umfragevariable besteht. Diese kann entweder die Folge davon sein, dass die Surveyvariable die Teilnahmewahrscheinlichkeit direkt beeinflusst (y beeinflusst p : in einer schriftlichen Studie zur Verbreitung des Analphabetismus (y) nehmen Analphabeten naturgemäß weniger teil (p)). Oder aber sowohl die interessierende Surveyvariable als auch die Teilnahmewahrscheinlichkeit werden beide durch eine dritte Variable beeinflusst (z als gemeinsame Ursache von y und p : in einer Wahlstudie beeinflusst das politische Interesse (z) sowohl die Umfrageteilnahme (p) als auch das Wahlverhalten (y)).

Die entscheidende Frage im Verhältnis von RR und NRB ist: Wie variiert der Unterschied zwischen Teilnehmern und Nichtteilnehmern mit Veränderungen in der RR? In der Vergangenheit wurde häufig (implizit) angenommen, dass der Unterschied zwischen Teilnehmern und Nichtteilnehmern mehr oder weniger fix ist. In diesem Fall würde jede Erhöhung der RR eine Reduzierung des NRB mit sich bringen. Empirische Studien, in denen die Effekte einer RR-Veränderung auf Umfrageergebnisse untersucht werden, zeigen, dass dem nicht so ist. Beispielsweise fanden amerikanische Forscher nur geringe Unterschiede in den Ergebnissen von Telefonumfragen, die aufgrund des unterschiedlichen Aufwandes in der Feldarbeit jeweils RR-Unterschiede von 25 Prozentpunkten aufwiesen (Keeter, Miller, Kohut, Groves & Presser, 2000; Keeter, Kennedy, Dimock, Best & Craighill, 2006). In einer vergleichbar angelegten Studie zum persönlich-mündlich durchgeführten ALLBUS 2008 war das Ergebnis ähnlich (Blohm & Koch, 2009). Trotz einer weitaus höheren RR in der begleitenden Zusatzstudie (63%) unterschieden sich deren Resultate nur bei wenigen Merkmalen von denen des regulären ALLBUS (mit 40% Ausschöpfung).

Aus diesen Ausführungen können mehrere Schlussfolgerungen gezogen werden.

1. Es ist nicht sinnvoll, ein bestimmtes (hohes) Level der Ausschöpfung zu definieren, bei dessen Erreichung oder Überschreitung das NR-Problem als gelöst einzustufen ist, weil kein NRB zu gewärtigen ist (sieht man von einer annähernd 100%igen RR ab). Im Gegenzug gibt es auch kein unteres Level der RR, ab der Umfrageergebnisse generell als problematisch (weil verzerrt) zu werten sind. Vorgaben, die beispielsweise die Akzeptanz eines Datensatzes alleine an die Erreichung eines bestimmten Ausschöpfungszieles knüpfen, entbehren deshalb der wissenschaftlichen Rechtfertigung.
2. NRB ist ein merkmalspezifisches Phänomen und sollte deshalb auch merkmalspezifisch bestimmt werden. Deshalb ist es auch kaum möglich, von dem NRB einer Umfrage zu sprechen. Am ehesten ist dies noch bei Umfragen denkbar, die auf ein Thema ausgerichtet sind. Weitaus komplizierter ist die Situation dagegen in multi-thematischen Umfragen, bei denen im Prinzip der NRB für dutzende oder gar hunderte von Merkmalen bestimmt werden kann.
3. Der in einer bestimmten Umfrage für verschiedene Merkmale beobachtbare NRB ist keine direkte Folge der gegebenen NRR per se, sondern vielmehr eine Konsequenz des jeweiligen Designs der Umfrage und der eingesetzten Prozeduren und Methoden (dem sog. 'survey protocol'). Die beiden Hauptquellen von NR (Nichterreichbarkeit und mangelnde Befragungsbereitschaft) etwa haben i.d.R. unterschiedliche Ursachen, und dementsprechend korrelieren sie häufig auch mit unterschiedlichen Surveyvariablen. Dementsprechend wird eine Umfrage andere Verzerrungen aufweisen, wenn die Ausfälle primär daraus resultieren, dass Zielpersonen nicht kontaktiert wurden, wie wenn die Ausfälle primär die Folge der mangelnden Kooperationsbereitschaft von Zielpersonen sind.

3. Methoden zur Bestimmung des Nonresponse bias

Will man der Frage des NRB empirisch nachgehen, steht man vor dem Problem, dass die beiden Komponenten, die über die Höhe des NRB entscheiden, in unterschiedlichem Maße einer Messung zugänglich sind: Während die NRR bzw. RR - zumindest theoretisch - gut und eindeutig ermittelbar ist, liegen Informationen über die zweite Komponente - den Unterschied zwischen Teilnehmern und Nichtteilnehmern bei einer Surveyvariable bzw. die Kovarianz zwischen der Surveyvariable und der Teilnahmewahrscheinlichkeit - normalerweise nicht vor. Es existieren jedoch verschiedene Methoden zur Annäherung an das Problem. Diese haben ihre spezifischen Vorzüge und Schwächen. Auch wenn nicht jede Methode bei jeder Umfrage einsetzbar ist, ist es ratsam, sich dem Problem mit verschiedenen

Ansätzen zu nähern, um ein besseres Verständnis der Situation in der betreffenden Umfrage zu erreichen. Die wichtigsten Methoden sind (vgl. Montaquila & Olson, 2012):

1. Vergleich der Umfrageergebnisse mit anderen Datenquellen / Aggregatstatistiken
2. Analyse von (Individual-)Daten für Teilnehmer und Nichtteilnehmer
3. Analyse von Variationen innerhalb der Gruppe der Umfrageteilnehmer
4. Analyse der Effekte unterschiedlicher Gewichtungsprozeduren

Vergleich der Umfrageergebnisse mit anderen Datenquellen / Aggregatstatistiken

Die am häufigsten verwendete Methode zum Studium des NRB ist der Vergleich von Umfragedaten mit den Ergebnissen einer anderen, genaueren Datenquelle. Ein Beispiel dafür ist der Vergleich von Umfrageergebnissen mit Daten des amtlichen Mikrozensus, welcher aufgrund der Teilnahmepflicht nur eine geringe NRR aufweist (vgl. z.B. Koch, 1998). Ein entsprechender Vergleich muss sich i.d.R. auf einige wenige soziodemographische Merkmale beschränken. Er setzt voraus, dass die Merkmale in vergleichbarer Form erhoben wurden, und dass die Grundgesamtheit und der Erhebungszeitpunkt zwischen der Umfrage und dem Mikrozensus nicht (stark) differieren. I.d.R. werden die Messungen allerdings nicht völlig identisch sein. Unterschiede können dann sowohl aus einem NR-Fehler als auch aus unterschiedlichen Messungen resultieren. Da beide, die Umfrage und der Mikrozensus, auf Stichproben basieren, ist überdies der Stichprobenfehler bei einem Vergleich zu berücksichtigen. Darauf hinzuweisen ist im Übrigen, dass aus dem Fehlen von Unterschieden in den betreffenden demographischen Merkmalen nicht auf die Unverzerrtheit der eigentlich interessierenden Surveymerkmale geschlossen werden kann. Ein Bias bei den demographischen Merkmalen bzw. dessen Fehlen sagt nur insoweit etwas über den Bias bei anderen Merkmalen aus, wie diese eng mit den betreffenden demographischen Merkmalen zusammenhängen. Peytcheva & Groves (2009) fanden in einer Metaanalyse, dass die Verzerrungen in den substantiellen Variablen nicht durch den Bias bei demographischen Merkmalen vorhergesagt werden konnten.

Analyse von (Individual-)Daten für Teilnehmer und Nichtteilnehmer

Eine weitere Methode NRB zu untersuchen nutzt Informationen, die für die gesamte Stichprobe (Teilnehmer und Nichtteilnehmer) einer Umfrage vorliegen. Wenn Informationen für die gesamte Stichprobe vorliegen, können die Daten für die Umfrageteilnehmer mit den Daten der Nichtteilnehmer verglichen werden, um Schätzungen für den NRB zu erlangen. Die Informationen können zum einen bereits Teil des Stichprobenrahmens selbst sein. Beispiele hierfür sind Informationen zu Alter und Geschlecht bei einer Einwohnermelderegisterstichprobe oder Informationen zur Dauer der Vereinsmitgliedschaft bei einer Befragung von Vereinsmitgliedern. Weitere Informationen können gewonnen werden, wenn Individualdaten aus anderen Quellen an die Stichprobe zugespielt werden können. In Betracht kommen hierfür etwa Daten aus administrativen Beständen (Gesundheitsdaten oder Daten der Beschäftigtenstatistik). Teilweise können auch Aggregatinformationen zu diesem Zweck verwendet werden. So enthalten beispielsweise die ursprünglich für Zwecke des Direktmarketing gesammelten Daten von MICROM (vgl. Microm, 2011) Informationen für Aggregate von acht Haushalten bzw. für Straßenabschnitte. Diese Informationen können auf der Basis von Haushaltsadressen an eine Stichprobe zugespielt werden (vgl. Goebel, Spieß, Witte & Gerstenberg, 2007). Eine weitere Quelle von Informationen für die gesamte Stichprobe stellen Beobachtungen der Interviewer dar. Im European Social Survey beispielsweise sollen die Interviewer für alle Fälle der Stichprobe den Haustyp klassifizieren, in dem die Zielperson lebt, sowie verschiedene Beobachtungen über die Wohnumgebung vornehmen und erfassen (Stoop, Billiet, Koch & Fitzgerald, 2010).

Der Vorzug dieser Methode ist, dass vergleichbare Messungen für Teilnehmer und Nichtteilnehmer der Umfrage vorliegen. Auf dieser Basis können für die betrachteten (Hilfs-) Merkmale Schätzungen des NRB erfolgen. Zusätzlich können Profile für die verschiedenen Ausfallgruppen (wie nichterreichte oder nichtkooperative Zielpersonen) erstellt und dadurch Hinweise auf die Quellen des NRB gewonnen werden. Allerdings stellen die vorliegenden Merkmale in der Regel nicht die eigentlich interessierenden Umfragevariablen dar (sonst wäre die Umfrage nicht notwendig gewesen!). Generell kommt es darauf an, dass die verwendeten Hilfsmerkmale („auxiliary variables“) eine enge Verbindung zu den eigentlich interessierenden Surveyvariablen haben. Wenn die Hilfsmerkmale eine enge Beziehung zu den Umfragevariablen aufweisen (was für die Gruppe der Umfrageteilnehmer empirisch untersucht werden kann), und die weiteren Analysen ergeben, dass sie auch mit dem Teilnahmeverhalten zusammenhängen, dann können daraus auch Hinweise über den vermutlichen NRB der eigentlich interessierenden Surveyvariablen gewonnen werden. Ein potentieller Nachteil der Methode ist, dass häufig nicht für alle Einheiten der Stichprobe die betreffenden Informationen verfügbar sind. Auch ist die Qualität der Messung der Hilfsvariablen teilweise fragwürdig. So variieren beispielsweise Beobachtungen häufig zwischen verschiedenen Interviewern (zur Qualität von Interviewerbeobachtungen vgl. z.B. Olson 2013).

Analyse von Variationen innerhalb der Gruppe der Umfrageteilnehmer

In diesem Design sollen durch den Vergleich verschiedener Subgruppen innerhalb der Umfrageteilnehmer Erkenntnisse über den NRB gewonnen werden. Dazu werden die Teilnehmer nach dem notwendigen Aufwand zur Erzielung eines Interviews („level of effort“) in eher „leichte“ oder eher „schwierige“ Fälle unterschieden. Die Aufteilung der Fälle kann sich dabei an unterschiedlichen Kriterien orientieren. So können z.B. (a) Befragte, mit denen das Interview bereits früh in der Feldzeit realisiert wurde, von denen unterschieden werden, die erst spät in der Feldzeit für eine Teilnahme gewonnen werden konnten. Oder Befragte können (b) nach der Anzahl der notwendigen Kontaktversuche differenziert werden. Denkbar ist auch (c) eine Unterscheidung danach, ob Befragte sofort zur Teilnahme bereit waren bzw. die Teilnahme zunächst abgelehnt haben und erst durch weitere Überzeugungsversuche für das Interview gewonnen werden konnten. Die für diese Klassifizierung notwendigen Informationen sind während der Feldarbeit, i.d.R. in Kontaktprotokollen, zu sammeln (sog. „paradata“). Der größte Vorteil dieser Methode ist, dass dann für alle Surveyvariablen untersucht werden kann, ob zwischen den definierten Gruppen Unterschiede auftreten. Der gravierendste Nachteil besteht darin, dass keine Informationen über die tatsächlichen Nichtteilnehmer vorliegen. Die – i.d.R. ungeprüfte – Annahme ist, dass die Nichtteilnehmer den „schwierigen“ Fällen ähnlicher sind als den „einfachen“ Fällen. Dies ist jedoch empirisch keineswegs zwingend der Fall, wie beispielsweise die Studie von Lin & Schaeffer (1995) zeigt. Eine weitere Beschränkung dieses Vorgehens liegt darin, dass häufig die Indikatoren der Schwierigkeit nicht eindeutig Prozessen der Erreichbarkeit bzw. der Befragungsbereitschaft zugeordnet werden können. Von Vorteil ist es beispielsweise, wenn Befragte nicht nach der Gesamtzahl der Kontaktversuche differenziert werden, sondern lediglich die Kontaktversuche bis zum ersten erfolgreichen Kontakt betrachtet werden, da dies eine klarere Zuordnung zum Problem der Erreichbarkeit erlaubt.

Ein Spezialfall der Analyse von Variationen innerhalb der Umfrageteilnehmer ist die Aufsetzung einer speziellen NR Follow-up Studie. Hier wird, i.d.R. im Anschluss an die eigentliche Umfrage, versucht, die Nichtteilnehmer (oder eine Substichprobe von ihnen) durch vermehrten Aufwand (z.B. durch monetäre Incentives) doch noch zur Umfrageteilnahme zu bewegen. Um die Erfolgchance zu erhöhen, wird dabei oft auch der Fragebogenumfang reduziert und/oder mit einem anderen Befragungsmodus gearbeitet (vgl. Stoop et al., 2010). Dies erschwert allerdings die Vergleichbarkeit zwischen den Teilnehmern der ursprünglichen Umfrage und den Teilnehmern der Follow-up Studie. Auch die unterschiedlichen Feldzeiten können in dieser Hinsicht Probleme aufwerfen, insbesondere dann, wenn nicht

ausgeschlossen werden kann, dass die gemessenen Merkmale sich verändern (wie z.B. bei der Erfassung von Einstellungen). Das grundsätzliche Problem von NR Follow-up Untersuchungen liegt jedoch darin, dass es trotz aller Anstrengungen normalerweise nicht gelingt mit allen Zielpersonen ein Interview durchzuführen – auch die Follow-up Studie ist dann weiterhin mit einem (mitunter erheblichen) NR-Problem konfrontiert.

Analyse der Effekte unterschiedlicher Gewichtungsprozeduren

Nur der Vollständigkeit halber sei hier noch erwähnt, dass durch den Vergleich von ungewichteten (bzw. lediglich design-gewichteten) Daten mit NR-gewichteten Daten Aufschluss darüber gewonnen werden kann, in welchem Umfang NRB durch die Gewichtung ausgeglichen werden kann (siehe auch GESIS Survey Guidelines Artikel „Gewichtung“ (Gabler, Kolb, Sans, & Zins, 2015)). Wenn unterschiedliche Gewichtungsprozeduren zur Verfügung stehen, können deren Ergebnisse untereinander und mit den ungewichteten Resultaten verglichen werden. Das Hauptproblem dieser Methode ist, dass man i.d.R. über keinen verbindlichen Standard verfügt, anhand dessen man bei beobachtbaren Ergebnisunterschieden entscheiden kann, welches Ergebnis die Realität besser wiedergibt.

4. Datenerhebung unter der Maxime der RR-Optimierung und NRB-Minimierung

Wer eine Umfrage plant, sollte sich aktiv mit dem NR-Problem auseinandersetzen. Üblicherweise wird man die Studie so anlegen, dass NR reduziert und eine möglichst hohe RR erzielt wird. Hierzu stehen verschiedene bewährte Maßnahmen zur Verfügung (vgl. Groves et al., 2009; Koch, Fitzgerald, Stoop, Widdop & Halbherr, 2012). In face-to-face Umfragen (genau wie in anderen Befragungsformen) wird man sich dabei nicht auf einzelne Maßnahmen beschränken, sondern ein ganzes Bündel von Maßnahmen einsetzen, um zu versuchen, Probleme der Erreichbarkeit, der Befragungsfähigkeit und der Kooperationsbereitschaft zu minimieren. Während manche Maßnahmen die Erfolgchancen in allen Teilprozessen (insbes. Kontaktierung und Motivierung) verbessern, sind andere spezifischer auf die erfolgreiche Kontaktierung oder die Motivierung von Zielpersonen gerichtet. So ist etwa der Einsatz erfahrener, gut geschulter und angemessen honorierter Interviewer generell eine gute Basis für eine erfolgreiche Studiendurchführung vor dem Hintergrund von NR. Die Festlegung einer ausreichend langen Feldzeit sowie Vorgaben zur Anzahl und Terminierung von Kontaktversuchen (Wochentag, Uhrzeit, zeitlicher Abstand, usw.) dagegen richten sich primär auf die erfolgreiche Kontaktierung von Zielpersonen. Im Gegenzug ist die Vermittlung von Strategien zur Vermeidung von Verweigerungen, der Einsatz von Ankündigungsschreiben oder von Incentives primär darauf angelegt, Zielpersonen zur Umfrageteilnahme zu motivieren.

Vor dem Hintergrund der obigen Ausführungen liegt es nahe nicht nur zu fragen, ob die genannten Maßnahmen helfen, die RR zu erhöhen, sondern auch inwiefern sie dafür geeignet sind, einen evtl. NRB zu vermeiden bzw. zu verringern. Darauf lässt sich kaum eine allgemein gültige Antwort geben. Die bisherige Forschung hat primär die Effekte der Maßnahmen auf die RR und kaum deren Effekte auf den NRB untersucht. Des Weiteren sollte man sich darüber im Klaren sein, dass die Wirkung einzelner Maßnahmen auch von den jeweiligen Gegebenheiten der spezifischen Umfrage abhängt, also von der untersuchten Population, dem Thema und der Länge einer Befragung, den weiteren Maßnahmen, die zur Erhöhung der RR geplant sind, usw. Technisch gesprochen dürften also eher Interaktions- und weniger Haupteffekte bei der Beurteilung der Wirksamkeit einzelner Maßnahmen zur Reduzierung von NRB im Fokus stehen (Groves, 2006; Groves & Peytcheva, 2008). Ein Beispiel: Wenn in einer Umfrage zu

einem Spezialthema ein Bias zu erwarten ist, weil Personen, die ein hohes Interesse an dem Thema haben, eine höhere Teilnahmewahrscheinlichkeit aufweisen, dann kann der Einsatz eines Befragtenincentives sinnvoll sein. Wie verschiedentlich gezeigt wurde, kann nämlich ein Incentive dazu beitragen, insbesondere Zielpersonen mit geringem Interesse am Befragungsthema zur Interviewteilnahme zu motivieren (Groves, Presser & Dipko, 2004). Allerdings ist auch nicht auszuschließen, dass in manchen Situationen ein Befragtenincentive kontraproduktiv ist, nämlich dann wenn es insbesondere die Teilnahmewahrscheinlichkeit von Personen erhöht, die bereits überproportional in der Stichprobe vertreten sind.

Groves' (2006, p. 668) generelle Empfehlung an den Umfragepraktiker lautet: „Blind pursuit of high response rates in probability samples is unwise; informed pursuit of high response rates is wise.“ In dieser Perspektive kann es also nicht darum gehen, unter allen Umständen eine Maximierung der RR anzustreben. Eine solche Maximierung geht häufig den Weg des geringsten Widerstandes und beinhaltet ein ‚go for the easy cases‘. Wenn zwischen den interessierenden Umfragevariablen und der Teilnahmewahrscheinlichkeit ein Zusammenhang besteht, dann kann durch eine solche Strategie ein vorliegender Bias sogar noch verstärkt werden (Beullens & Loosveldt, 2012).

Ganz allgemein formuliert könnte die Zielvorstellung zur NRB-Minimierung bei der Datenerhebung darin bestehen, am Ende der Feldarbeit für alle Stichprobeneinheiten die gleiche Teilnahmewahrscheinlichkeit zu erzielen. Dies erfordert, Zielpersonen mit niedriger Teilnahmewahrscheinlichkeit zu identifizieren und diese mit besonderer Intensität zu bearbeiten (Peytchev, Riley, Rosen, Murphy & Lindblad, 2010). Hierfür sind Hilfsvariablen notwendig, die sowohl einen Einfluss auf die Teilnahmewahrscheinlichkeit haben als auch mit den Surveyvariablen in engem Zusammenhang stehen. Die Identifizierung und Sammlung entsprechender Informationen ist kein einfaches Unterfangen. Häufig stehen nur einige wenige demographische Daten (wie Alter oder Geschlecht der Zielperson aus dem Stichprobenrahmen oder Beobachtungen der Interviewer zur Wohnung und Wohnumgebung) zur Verfügung. Wenn gemäß diesen Hilfsvariablen bestimmte Gruppen eine zu niedrige RR aufweisen, sind gezielte Interventionen einzuleiten, um die Befragungsbemühungen für diese Gruppe zu intensivieren. Dies kann die Erhöhung der Kontaktbemühungen beinhalten, den Einsatz besonders kompetenter Interviewer oder das Angebot eines (höheren) monetären Incentives.

Die Implementierung eines entsprechenden Systems zielgerichteter Interventionen erfordert die intensive Beobachtung und Analyse von Prozessen der Feldarbeit sowie die Fähigkeit, tatsächlich auch steuernd eingreifen zu können (zu sog. ‚responsive designs‘ siehe Groves & Heeringa, 2006). Wenn zur Realisierung einer face-to-face Bevölkerungsumfrage mit einem kommerziellen Umfrageinstitut kooperiert wird, sollte nicht unterschätzt werden, dass der Installierung dieser Vorgehensweise erhebliche Schwierigkeiten entgegenstehen. Die vorherrschende Philosophie und die organisatorischen Abläufe (wie Interviewerschulungen oder -bezahlung) in der Umfragepraxis sind primär auf die Erzielung einer bestimmten Fallzahl von Interviews und einer möglichst hohen RR ausgerichtet. Interviewer sind als Freiberufler tätig, arbeiten oft für mehrere Institute und sind kaum weisungsgebunden. Eine Orientierung der Feldarbeit auf bestimmte – häufig nur überproportional schwer zu erreichende oder zu motivierende – Zielpersonen ist in dieser Umgebung nur bedingt umzusetzen. Unabhängig von diesen Hindernissen der Implementation ist im Übrigen nochmals zu betonen, dass der Erfolg eines entsprechenden Vorgehens damit steht und fällt, wie es gelingt, gute Hilfsvariablen zu finden. Wenn die verfügbaren Hilfsvariablen zwar die Teilnahmebereitschaft beeinflussen, aber nur wenig mit den zentralen Umfragevariablen zusammenhängen, kann eine Ausrichtung der Feldarbeit an ihnen zwar (mitunter erheblichen) Mehraufwand bedeuten, ohne dass dadurch jedoch der NRB nennenswert reduziert würde.

5. Schlussbemerkung

In dem Zitat, das dem Beitrag vorangestellt ist, wird Nonresponse als das am häufigsten missverstandene und umstrittenste Thema der Umfrageforschung bezeichnet. Wenig kontrovers dürfte sein, dass NR auch zukünftig eine der zentralen Herausforderungen der Umfrageforschung darstellt. Es wird kaum gelingen, die vorherrschenden niedrigen RR von Umfragen erheblich zu steigern, und auch die Ansätze, NRB bereits in der Feldarbeit zu minimieren, werden keine umfassende Lösung des Problems bieten. Vor diesem Hintergrund wird die Berücksichtigung des NR-Problems durch statistische Korrekturverfahren während der Datenanalyse an Bedeutung gewinnen (Gabler et al., 2014; Brick, 2013). Auch in diesen Ansätzen ist die Verfügbarkeit bzw. Gewinnung von aussagekräftigen Hilfsvariablen von zentraler Bedeutung. Die wissenschaftliche Durchdringung des Teilnahmeverhaltens an Umfragen sowie die planvolle Gestaltung des gesamten Datenerhebungsprozesses stehen damit auf der Agenda.

Literaturverzeichnis

- Beullens, K., & Loosveldt, G. (2012). Should high response rates really be a primary objective? *Survey Practice*, 5(3), 1–5.
- Blohm, M., & Koch, A. (2009). *Ausschöpfungsquoten und Stichprobenqualität am Beispiel des ALLBUS 2008: Führt eine höhere Ausschöpfung zu anderen/besseren Umfrageergebnissen?* Vortrag beim Workshop "Qualität von Großerhebungen". Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, 3.–4. Sept. 2009.
- Brick, M. J. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29(3), 329–353.
- Dixon, J., & Tucker, C. (2010). Survey nonresponse. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (2nd ed.) (pp. 593–630). Bingley: Emerald.
- Gabler, Siegfried, Kolb, Jan-Philipp, Sand, Matthias & Zins, Stefan (2015). Gewichtung. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/sdm-sg_009
- Goebel, J., Spieß, C. K., Witte, N. R. J., & Gerstenberg, S. (2007). *Die Verknüpfung des SOEP mit MICROM-Indikatoren: Der MICROM-SOEP Datensatz*. (DIW Data Documentation 26). Berlin: Retrieved from http://www.diw.de/documents/publikationen/73/diw_01.c.78103.de/diw_datadoc_2007-026.pdf
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646–675.
- Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.
- Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A meta-analysis. *Public Opinion Quarterly*, 72(2), 167–189.
- Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68(1), 2–31.

- Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3), 439-457.
- Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, 64(2), 125-148.
- Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly*, 70(5), 759-779.
- Koch, A. (1998). Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. *ZUMA-Nachrichten*, 42, 66-90.
- Koch, A., Fitzgerald, R., Stoop, I., Widdop, S., & Halbherr, V. (2012). *Field procedures in the European Social Survey round 6: enhancing response rates*. Retrieved from http://www.europeansocialsurvey.org/docs/round6/methods/ESS6_response_enhancement_guidelines.pdf
- Lin, I.-F., & Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly*, 59(2), 236-258.
- Microm (2011). *Microm Datenhandbuch. Arbeitsunterlagen für microm MARKET & GEO*. Neuss: microm Micromarketing-Systeme und Consult GmbH.
- Montaquila, J. M., & Olson, K. M. (2012). Practical tools for nonresponse bias studies. SRMS/AAPOR Webinar. Retrieved from <http://www.amstat.org/sections/srms/webinarfiles/NRBiasWebinarApril2012.pdf>
- Olson, K. M. (2013). Paradata for nonresponse adjustment. *The Annals of the American Academy of Political and Social Science*, 645, 142-170.
- Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *Survey Research Methods*, 4(1), 21-29.
- Peytcheva, E., & Groves, R. M. (2009). Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *Journal of Official Statistics*, 25(2), 193-201
- Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving Survey Response: Lessons Learned from the European Social Survey*. Chichester: Wiley.

Weitere Literaturhinweise:

- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Schnell, R. (1997). Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske & Budrich. Retrieved from http://kops.ub.uni-konstanz.de/xmlui/bitstream/handle/urn:nbn:de:bsz:352-opus-56148/Nonresponse_in_Bevoelkerungsumfragen.pdf?sequence=1

Sonderhefte von Zeitschriften zum Thema ‚Nonresponse‘:

- The Annals of the American Academy of Political and Social Science* (2013), 645(1): The Nonresponse Challenge to Surveys and Statistics
- Journal of Official Statistics* (2011), 27(2)

Journal of the Royal Statistical Society: Series A (2013), 176(1): The use of paradata in social survey research

Public Opinion Quarterly (2006), 70(5): Special Issue: Nonresponse Bias in Household Surveys